Deep learning - 1. Assignment - Basic Neural Network

Nejc Ločičnik (63180183)

1 Introduction

This assignment is about convolutional neural networks (CNN). Main focus is on the ResNet18 (+FCN) and U-Net architectures. The task is split into 3 tasks based on the problem we are solving: classification, semantic segmentation and colorization.

2 Classification with ResNet18

The first task is about bird classification using the ResNet18 backbone architecture. The results of experiments and their parameters are represented in table 1. I started with a base network, then added regularization, increased batch size, amount of epochs. The final experiment uses step learning rate schedule, which boosted accuracy the most.

Learn. r.	Epochs	Batch	Dropout	W. decay	Accuracy
0.001	10	32	/	/	0.89
0.001	15	64	/	/	0.91
0.001	25	128	0.5	0.0001	0.91
0.001	25	256	0.5	0.0001	0.94

Table 1: Result of tuning parameters (last one uses step learning rate schedule).

The fails (figure 1) are mostly on similar birds, usually with a very distinct color scheme or very similar shape. Most of the failures are on different subspecies, so I think the results are pretty good. We'd need a deeper network (to abstract out more specific features) and more training to further raise the accuracy.

Figure 1: Examples of failed classifications.

3 Semantic segmentation

For semantic segmentation we compare results of ResNet18 (adapted with FCN in the last layer) and U-Net. Both networks used the parameters from task 1 with less epochs (10) and the exception of U-Net not having dropout and batch normalization. The results are evaluated using mean IoU, which I calculated by dividing the sum of correctly classified pixels with the image size for each sample. This was averaged out for the whole test set. The results are presented in table 2.

Examples of the predicted segmentation for both networks is presented in figures 2 and 3. We can see that U-Net retained more specific object shape compared to ResNet. This is because of its structure, U-net can be split in two parts: encoder and decoder, encoder acts like a general CNN, where spatial dimensions are gradually reduced, abstract different levels of features (encoding feature information), while decoder gradually upsamples the obtained features back to the original image size. The encoder and decoder are additionally connected with skip-connections in each layer of downsampling/upsampling, which helps produce high-resolution segmentation maps. The importance of skip-connections in U-Net are further explained in section 4.

ResNet on the other hand only abstracts out features (like the encoder in U-Net) and then directly upsamples them to the original image size where each pixel is classified (FCN adaptation). The lack of skip-connections and upsampling means

Network	Mean IoU
ResNet18+FCN	92.103%
U-Net	94.703%

Table 2: Comparison of IoU results between ResNet18+FCN and U-Net.

there is a lacking connection between the obtained features and specific pixels in the original image, which results in a more general shape in the segmentation map. This is most notable in how tree canopies are segmented.

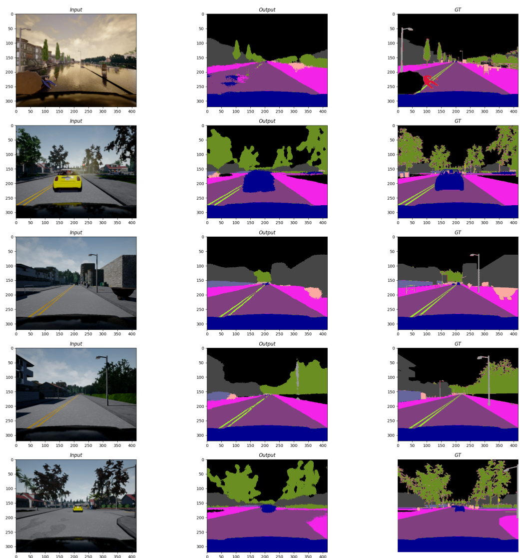

Figure 2: Result of segmentation with ResNet18+FCN.

Figure 3: Result of segmentation with U-Net.

4 Colorization

The last task is about colorizing gray scale images. Here we are looking at the difference between a U-Net with and without skip-connections between the encoder and decoder. Results are presented in figures 4 and 5.

We can see that not having skip-connections results in a blurry image. This is because of the bottleneck between the encoder and decoder. We lose a lot of detailed information when encoding features, which is then not retained in upsampling wihtout the skip-connections.

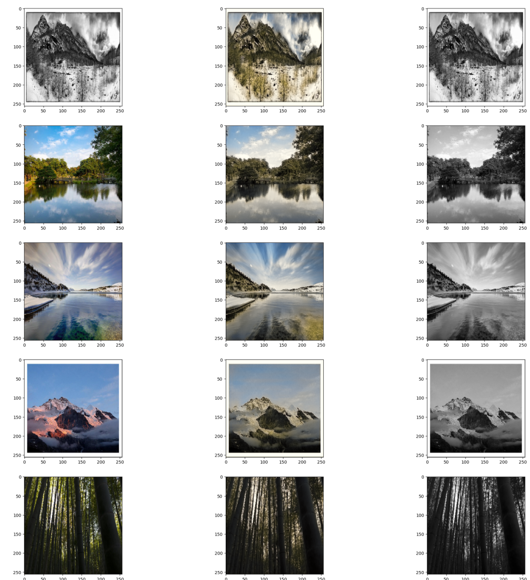

Figure 4: Result of colorization with U-Net with skip-connections.

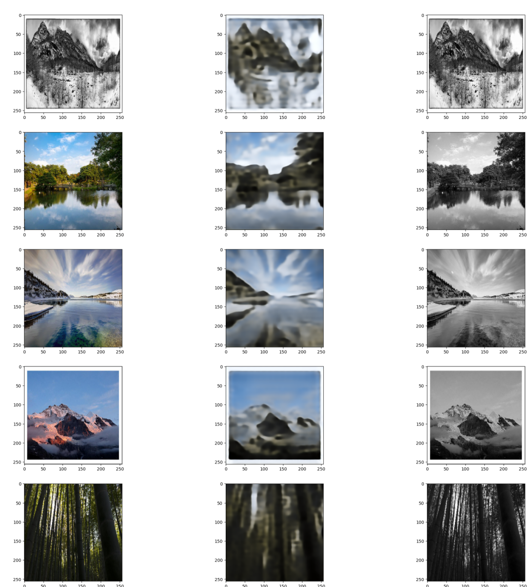

Figure 5: Result of colorization without U-Net with skip-connections.

As we can see we don't fully colorize the images. The results look like overlapping the gray-scale image with a transparent (alpha=0.5) RGB image. I'm not fully sure if this is because of the skip connections (as the input is a gray-scale image) or because some ground truths are also gray-scale images. The second thing I noticed is that red isn't colored as much. I think this might be because warm colors are underrepresented in the dataset, which is about landscapes (so a lot of white/blue/green/gray and less red/warm colors).

Training a model on each channel (R, G, B) and then combining the results might produce different results. I didn't have time to try it.

5 Conclusion

In this assignment, we explored the effectiveness of convolutional neural networks (CNNs) in various tasks: classification, semantic segmentation, and colorization. We utilized the ResNet18 architecture for bird classification, achieving improved accuracy by tuning parameters and implementing a step learning rate schedule. For semantic segmentation, we compared ResNet18+FCN and U-Net, finding that U-Net outperformed due to its encoder-decoder structure with skip connections. Lastly, in colorization, we observed that skip connections in U-Net play a crucial role in preserving image details during upsampling. Overall, CNNs demonstrated promising capabilities across diverse image processing tasks, with architecture selection and parameter tuning playing key roles in achieving optimal performance.